ABSTRACT

Several cluster chips and a shared main memory are connected by interconnect buses. Each cluster chip has multiple processors using multiple level-2 local caches, two memory controllers and two snoop tag partitions. The interconnect buses connect all local caches to all snoop tag partitions on all cluster chips. Each snoop tag partition has all the system's snoop tags for a partition of the main memory space. The snoop index is a subset of the cache index, with remaining chip-select and interleave address bits selecting which of the snoop tag partitions on the multiple cluster chips stores snoop tags for that address. The number of snoop entries in a snoop set is equal to a total number of cache entries in one cache index for all local caches on all cluster chips. Cache coherency request processing is distributed among the snoop tag partitions on different cluster chips, reducing bottlenecks.

THIS PAGE BLANK (USPTO)